Efficacy of a High School Extensive Reading Intervention for English Learners With Reading Difficulties

Sharon Vaughn, Leticia R. Martinez, Kelly J. Williams

The University of Texas at Austin

Jeremy Miciak

The University of Houston

Anna-Maria Fall and Greg Roberts

The University of Texas at Austin

Journal of Educational Psychology 2018

Sharon Vaughn, Leticia R. Martinez, and Kelly J. Williams, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin; Jeremy Miciak, Psychology Department, The University of Houston; Anna-Mária Fall and Greg Roberts, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin.

Kelly J. Williams is now at Department of Curriculum and Instruction, Indiana University.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150058 to The University of Texas at Austin. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Sharon Vaughn, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1912 Speedway, SZB 228, Austin, TX 78712. E-mail: srvaughn@austin.utexas.edu

Abstract

This study examined the effects of Reading Intervention for Adolescents, a 2-year extensive reading intervention targeting current and former English learners identified as struggling readers based on their performance on the state accountability assessment. Students who enrolled at three participating urban high schools were randomly assigned to the Reading Intervention for Adolescents treatment condition (n = 175) or a business-as-usual comparison condition. Students assigned to the treatment condition participated in the intervention for approximately 50 minutes daily for 2 school years in lieu of a school-provided elective course, which business-as-usual students took consistent with typical scheduling. Findings revealed significant effects for the treatment condition on sentence-level fluency and comprehension (g = 0.18) and on a proximal measure of vocabulary learning (g = .41), but not on standardized measures of word reading, vocabulary, or reading comprehension (g range: -0.09 to 0.06). Posthoc moderation analyses investigated whether initial proficiency levels interacted with treatment effects. On sentence-level fluency and comprehension and on vocabulary learning, initial scores were significantly associated with treatment effects—however, in opposite directions. Students who scored low at baseline on sentence reading and comprehension scored relatively higher at posttest on that measure, whereas students who scored high at baseline on the proximal vocabulary measure scored relatively higher at posttest on that measure. The discussion focuses on the difficulty of remediating persistent reading difficulties in high school, particularly among English learners, who are often still in the process of acquiring academic proficiency in English.

Keywords: English learners, high school, reading difficulties, intensive reading, intervention

Efficacy of a High School Extensive Reading Intervention for English Learners With Reading Difficulties

High school English learners (ELs) with significant reading difficulties may represent one of the most at-risk subgroups in American schools. Their heightened risk status reflects the severe consequences of low reading ability and their limited remaining time in school, time during which they might receive formal instruction to remedy their reading problem. Reading difficulties in students are associated with a myriad of deleterious outcomes, including an increased likelihood that they fail to complete high school on time or drop out altogether (Hernandez, 2011), lower rates of participation in the work force (National Research Council, 2011), and overall lower earnings than individuals who complete high school (Snyder, de Brey, & Dillow, 2016). By definition, ELs in high school demonstrate relatively low language proficiency, which manifests in difficulty reading, writing, and speaking in English and affects their successful participation in school and postsecondary opportunities. Though ELs often participate in targeted programs to facilitate their English language development, many continue to demonstrate significant reading deficits. ELs perform significantly lower than non-ELs on the National Assessment of Educational Progress reading assessment; in 2016, 67% of 8th-graders and 73% of 12th-graders scored below the basic level of proficiency (National Center for Education Statistics, 2016). ELs are more likely than non-ELs to drop out of school (Hernandez, 2011; Kennelly & Monrad, 2007). One urban school district has reported that ELs drop out at a 43% rate, whereas non-ELs drop out at a 32.5% rate. Furthermore, ELs re-enroll at a lower rate (25.6%; Berliner, Barrat, Fong, & Shirk, 2008).

Challenges of Developing Reading for High School Students

Despite a robust literature documenting the effectiveness of early reading interventions for improving word reading outcomes in early grades (Gersten et al., 2008; Wanzek et al., 2016; Wanzek & Vaughn, 2007), many middle grade students with reading difficulties demonstrate inadequate word reading and reading comprehension. Furthermore, 81% of middle grade students with reading difficulties demonstrate problems in not only reading comprehension, but also fluency and decoding (Cirino et al., 2013). Brasseur-Hock et al. (2011) reported that in a sample of adolescents with reading difficulties, those who scored lowest on a reading comprehension assessment were most likely to have reading difficulties across all reading component areas, including decoding, fluency, and comprehension.

What about interventions for older students with reading difficulties? Considerably more is known about students in grades 4 through 8 than students in grades 9 through 12 (Wanzek, Vaughn, Scammacca et al., 2013; Wanzek, Wexler, Vaughn, & Ciullo, 2010). Syntheses examining the effects of interventions report either no intervention studies or very few intervention studies addressing high school students with reading disabilities (Edmonds et al., 2009; Scammacca, Roberts et al., 2016; Wanzek, Vaughn, Scammacca et al., 2013). Studies investigating the efficacy of interventions for improving literacy outcomes for students with reading difficulties in grades 4 through 9 yield consistent results; most studies document no impact or low impact on reading comprehension (Flynn, Zheng, & Swanson, 2012; James-Burdumy et al., 2012; Scammacca, Roberts et al., 2016). For example, results from randomized control trials (RCTs) investigating yearlong interventions provided to struggling readers in middle school have yielded small standardized effects ranging from ES = 0.09 to ES = 0.16 (Corrin, Somers, Kemple, Nelson, & Sepanik, 2008; Kemple, Corrin, Nelson, & Salinger, 2008),

regardless of whether the intervention focused primarily on text reading and comprehension (Vaughn, Klingner et al., 2011), was aligned with the reading skills of the target students (Vaughn, Wexler et al., 2011), or included technology components (Lang et al., 2009). The U.S. Department of Education funded 17 RCTs to examine the effects of interventions on students' reading comprehension outcomes, yielding a mean effect of ES = 0.11 (Boulay, Goodson, Frye, Blocklin, & Price, 2015). Of the 10 interventions investigated (some interventions had multiple RCTs), only 4 had positive or mixed results and the rest were not associated with any discernable effects.

Over the past 10 years, the research team has conducted more than 12 efficacy studies with students with reading difficulties in the middle grades. The studies have systematically investigated questions related to dosage, duration, delivery, and foci through (a) variations in group size (Vaughn, Wanzek et al., 2010; Wanzek, Vaughn, Roberts, & Fletcher, 2011); (b) variations in time allocated to intervention components, such as fluency, comprehension, and word study (Vaughn et al., 2000; Wexler, Vaughn, Roberts, & Denton, 2010); (c) a comparison of standardized versus more customized interventions (Vaughn, Wexler et al., 2011); and (d) a study of the impact of extensive interventions provided over multiple years (Vaughn et al., 2012; Vaughn, Roberts, Wexler et al., 2015; Vaughn et al., 2016).

One study examined the long-term effects of a 3-year treatment with students with intractable reading difficulties that was implemented in sixth grade and continued through eighth grade (Vaughn, Wexler, Leroux et al., 2012). Results from the 3-year study yielded an impressive effect size on a standardized measure of reading comprehension, ES = 1.20. However, this robust impact on reading comprehension was a function of both decline on the part of the comparison students and growth on the part of the treatment students.

The previous study that most resembles the present study is an RCT that was conducted with struggling readers in 9th and 10th grades. It compared the reading performance of students assigned to a 2-year reading intervention or to a business-as-usual (BaU) condition, which was an elective high school class. Findings revealed an impact of ES = 0.44 on a standardized reading comprehension measure following the 2-year intervention (Vaughn, Roberts, Wexler et al., 2015).

The most consistent finding to emerge from this work investigating reading interventions with secondary students has been in support of extensive interventions (i.e., interventions lasting more than 1 year) having a positive impact on reading comprehension for students with intractable reading problems. Despite these encouraging findings, none of the previous work investigating reading interventions in secondary settings systematically examined ELs, the largest growing population of students.

ELs With Reading Difficulties

ELs represent a diverse population of students with a broad variety of strengths and needs, including varying degrees of language proficiency, in their native languages and in English, and varying degrees of subject matter knowledge. Within the next 10 to 15 years, it is expected that as many as one in four children enrolled in U.S. schools will be ELs, and many of them will face significant challenges in reading comprehension (Lesaux, Kieffer, Kelley, & Harris, 2014; Mancilla-Martinez, Kieffer, Biancarosa, Christodoulou, & Snow, 2011). Researchers have claimed that a single label (EL) belies the diversity of the population (August & Shanahan, 2006). Genesee, Lindholm-Leary, Saunders, and Christian (2006) identified numerous sources of heterogeneity among ELs, including individuals' approaches to learning, the extent of transfer from their home language to English, and the influence of their home

language on school achievement. However, Kieffer (2008; 2011) found noteworthy differences among subpopulations of ELs when contrasting reading growth trajectories of limited English proficiency (LEP) students with those of English-only students from kindergarten to fifth or eighth grades. The average reading growth trajectory of language-minority students who entered kindergarten with adequate English proficiency was indistinguishable from that of their English-only peers; in contrast, language-minority students who entered school with the LEP designation grew on average more slowly in their reading scores than their English-only counterparts, with the reading achievement gap between these two groups increasing throughout elementary and middle school. Though there are many similarities in the way that ELs and non-ELs express reading comprehension difficulties, there are also differences to consider, such as the consistently low vocabulary reported for ELs (Lesaux & Kieffer, 2010) and their need for additional oral language supports (Lesaux & Harris, 2017).

Research has established a pattern of significant decline in reading over time for ELs with reading difficulties. Yet, with reviews indicating few interventions and research-based instructional practices designed to meet their needs, it is challenging to identify instructional practices associated with improved reading outcomes for these students (Short & Fitzsimmons, 2007). A practice guide published by the Institute of Education Sciences synthesizing research on reading interventions for ELs did not include ELs in high school after determining that there was so little research on that group (Baker et al., 2014). Findings from a meta-analysis of 11 eligible studies summarizing reading interventions for ELs with reading difficulties in grades 4 through 8 (Hall, Roberts, Cho et al., 2017) revealed that the mean meta-analytic effect size for all measures was moderate ($g = 0.35$). However, when standardized reading measures were examined separately, the mean meta-analytic effect size was near 0 ($g = 0.01$). Taken together,

the lack of rigorous research in this area and the limited effectiveness of interventions studied highlight a critical need to develop and examine the effects of rigorous interventions that integrate best practices in reading instruction for high school ELs with persistent reading problems. Based on this need, we conducted a 2-year RCT with students in 9th through 10th grades to determine the efficacy of a reading intervention, Reading Intervention for Adolescents (RIA), which was provided to students in place of an elective course.

Reading Intervention for Adolescents

The efficacy of RIA has been evaluated across multiple studies at the middle and high school level (Roberts, Vaughn, Fletcher, Stuebing, & Barth, 2013; Vaughn, Roberts, Schnakenberg et al., 2015; Vaughn, Roberts, Wexler et al., 2015). RIA consists of two instructional phases: Phase I emphasizes word study, fluency, vocabulary, and comprehension at the sentence and paragraph levels; Phase II extends Phase I instruction and emphasizes vocabulary and comprehension within content area texts. In the current study, Phase I was implemented in the first semester of year 1, and Phase II was implemented in the second semester of year 1 and continued through all of year 2.

The impact of the Reading Excellence: Word Attack and Rate Development Strategies (REWARDS) program, which was implemented in Phase I, has been examined as part of multicomponent reading interventions and in standalone evaluations (Kundert, Newman, Gifford, Haase, & Clure, 2012; Shippen, Houchins, Steventon, & Sartor, 2005; Vaughn et al., 2010; Vaughn, Roberts, Schnakenberg et al., 2015). Vaughn and colleagues (2010; 2015) used REWARDS Intermediate and REWARDS Secondary as part of multicomponent reading interventions for struggling readers in the secondary grades. Although it is not possible to discern the impact of the REWARDS program alone, students who participated in these interventions

made significant growth on measures of word attack, spelling, and passage comprehension (Vaughn et al., 2010; Vaughn, Roberts, Schnakenberg et al., 2015).

Phases I and II also included explicit instruction in comprehension strategies through Collaborative Strategic Reading (CSR), which is a text-based collaborative learning approach (Klingner & Vaughn, 1998). In a cluster randomized trial, middle and high school science and social studies teachers were trained to implement CSR, and students who used CSR in these content area classes at least once a week had significantly higher reading comprehension outcomes than students who did not receive the CSR instruction (Boardman, Klingner, Buckley, Annamma, & Lasser, 2015). In an analysis of two separate studies of CSR, Boardman and colleagues (2016) reported that higher quality CSR instruction was associated with higher comprehension scores for students with disabilities.

RIA with ELs. In the current study, RIA was adapted to meet the needs of ELs. The instructional enhancements were based on several syntheses and meta-analyses of effective instruction for ELs (August & Shanahan, 2006; Baker et al., 2014; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006). In each unit, a high-quality, engaging informational text was first selected, and then a small number of high-utility academic vocabulary words that related to the main idea of the text were explicitly taught (Baker et al., 2014; Francis et al., 2006). Students received a graphic organizer with each of the words in the unit, along with student-friendly definitions from the Merriam-Webster Learner's Dictionary, synonyms, antonyms, sample sentences, and visual examples and nonexamples of the words. Students were provided multiple opportunities to interact with these words in writing, speaking, and listening activities. Additionally, in CSR, students were taught independent word-learning strategies to help them determine the meaning of unknown words. Students had to reread the sentence with the unknown

word, reread the sentence before and after the unknown word, break the word into meaningful parts (i.e., prefixes, suffixes, roots), and/or use their native language to identify cognates.

In addition to explicit vocabulary instruction, RIA was enhanced for ELs to provide many opportunities to use oral and written academic language across the content areas (Baker et al., 2014; Lesaux & Harris, 2017). Each unit included a short video, visuals, and graphic organizers designed to increase student understanding. Students were placed in collaborative groups, which allowed them daily opportunities to discuss content. Structured activities such as turn and talk provided students with prompts to use their newly learned academic vocabulary words in context with their peers. Students also completed writing and oral activities that extended their learning by requiring them to produce sentences, summaries, and book share presentations.

Purpose of the Study and Research Questions

The purpose of this study was to examine the effects of a 2-year extensive reading intervention, RIA, on the reading achievement of current and former ELs identified as struggling readers in high school based on performance on the reading portion of their eighth-grade state accountability test. Specifically, we addressed the following research questions: 1. What are the effects of an extensive reading intervention on the reading achievement of ELs who are struggling readers compared with a school-implemented comparison group? 2. Does initial status on reading outcomes moderate the effects of the reading intervention?

Method

This RCT examined the effects of RIA on reading outcomes for 9th- and 10th-grade ELs with reading difficulties. Eligible participants (n = 358) were blocked by school and randomly assigned to one of four conditions: RIA only, dropout prevention intervention (DO) only, RIA and DO (RIA+DO), or BaU. Participants remained in the same condition throughout their 9th

and 10th grade years. For the purpose of this analysis, two distinct groups (i.e., treatment and comparison) were formed from the four groups that were originally randomized. The treatment group consisted of participants assigned to RIA only and RIA+DO, and the comparison group consisted of participants assigned to DO only and BaU. This represents a partially nested design with treatment students nested in reading tutor. There is no nesting of comparison students in tutor. We ignored clustering effects because the intraclass correlations across tutors (and across schools, with which tutors is confounded) were trivial (<.05). We assumed that the DO intervention, a modified version of Check & Connect (University of Minnesota, 2015), does not affect reading outcomes (or that its effect is non-different across the groups), and we tested this assumption by comparing scores on reading outcomes for students in the DO condition with those in the BaU condition. There were no significant differences (*p*-values ranged from .26 to .85), confirming our assumption that it was reasonable to combine groups.

Participants

School sites. Participants were recruited from a diverse, urban school district in the southwestern United States. Administrators selected three large high schools in the participating district with many ELs or students identified as ELs within the past 5 years. Demographic data for each school were obtained from the state educational agency website for the 2015–2016 school year, the first year of the study. Across schools, the majority of students were Hispanic (54.9%–90.5%) and economically disadvantaged (75%–90%) based on free and/or reduced-price lunch status. The percentage of currently identified ELs in the schools ranged from 13.4% to 49.4%. Two schools met state accountability standards; the third school had a "Needs Improvement" status due to poor performance on the state's school accountability measures.

Participants. Students needed to be screened for eligibility prior to entering high school so that students randomized to the intervention condition it could be assigned to the reading intervention class. Class scheduling typically occurs in the summer after eighth-grade and prior to students attending their ninth-grade year in high school and participating schools asked that we randomize and schedule students in the summer to avoid large changes to schedules in the fall. Eighth-grade students who were projected to attend one of the three participating high schools were screened for participation in the study. Eligible participants met the following two criteria: (a) a school designation of LEP or an LEP designation in the previous 5 years and (b) a score of 1,612 or below on the eighth-grade reading portion of the State of Texas Assessments of Academic Readiness (STAAR). We selected this cut point for eligibility because a score of 1,612 is one standard error above the failing score of 1,575, and we wanted to account for inherent measurement error and recognized that students close to the failing score of 1,575 on both sides will present highly similar reading profiles and may need continued interventions. The STAAR measures students' understanding of literary and informational text through 44 multiple-choice, text-based questions (Texas Education Agency, 2016). ELs who had been in the United States for less than a year were excluded from participation, as these students often have very limited English proficiency and RIA was not designed to address their needs.

At the beginning of the intervention, the sample included 358 participants, including 226 currently identified ELs and 132 formerly identified ELs (see Table 1). The majority were male (52.5%, $n = 188$), Hispanic (89.4%, $n = 320$), and economically disadvantaged (75.4%, $n = 270$). In addition, 44 students were receiving special education services, and the majority of them ($n = 41$) were identified as having a learning disability.

Attrition analyses. The program was implemented in the fall of 2015. To accommodate class scheduling, which occurred during the summer of 2015, we identified and randomized eligible eighth-grade students in the spring of 2015. This plan required that we draw our sample from the middle schools that typically "feed" the high schools where we had permission to implement the program. However, because the participating district's students are permitted to attend the high school of their choosing, we expected that only a subset of the recruited and randomized students would enroll at one of the targeted high schools in the fall of 2015. We randomized 713 eighth-grade students to one of four conditions within the high school they were most likely to attend. As a result of families being able to choose their high school, about half of the randomized sample did *not* enroll at one of the participating high schools in the fall (n = 355). Of the 358 students who enrolled, 318 completed the posttest battery at the end of ninth grade, which represents 44.6% of the original randomized sample and 89% of the Time 1 sample. At the beginning of 10th grade, 306 students completed the year 2 pretest battery (42.92% of the randomized sample and 85.47% of the Time 1 sample). A total of 264 students (37.03% of the randomized sample and 73.74% of the Time 1 sample) had complete data at the end of 10th grade.

Because students in 8th grade have options about where they attend high school (not required to attend the high school from their middle school feeder pattern), we expected overall attrition from eighth grade to ninth grade to be about 50% based on past studies (Vaughn, Roberts, Wexler et al., 2015) involving cohorts transitioning to high school in large districts that permit school choice. We oversampled accordingly, and the study is adequately powered to detect treatment effects on reading achievement outcomes that are typical for 2-year reading interventions with older, struggling readers. Nonetheless, the high rate of overall attrition

represents a threat to the findings' internal validity, power considerations aside. To the extent that students' attrition is related to group assignment in any way, an overall rate of 63% represents a legitimate threat to the findings' integrity, particularly given that the reading program was paired with a dropout prevention program (What Works Clearinghouse, 2017). However, the bulk of attrition occurred during the summer before implementation because students decided to attend schools other than those in the school-level sample, a decision we assume was unrelated to the program's pending implementation because we had yet to pursue parent and student consent to participate in the project (i.e., the parents and students were largely unaware of the program). For these reasons, we apply the What Works Clearinghouse's liberal attrition standard to evaluate the bias associated with overall and differential attrition. Table 2 summarizes these values across waves of the study. The combination of 63% overall attrition and 5% differential attrition at Time 4 is associated with tolerable amounts of attrition.

Baseline equivalence. In Table 3, we summarize the two groups' comparability at baseline. Baseline equivalence was established using preintervention test scores for all outcome measures of interest. Per the What Works Clearinghouse's recommendations (2017), we report standardized differences for groups' pretest means. All differences are within tolerable limits (ES < .26), except for the pretest difference on the Test of Word Reading Efficiency–Second Edition (TOWRE-2) sight word subtest which was ES = .26.

Reading Intervention for Adolescents

Students in the treatment conditions (RIA only and RIA+DO) received the RIA class for 2 full school years in grades 9 and 10 (2015–2017). Students were taught in groups of 10 to 15 students, as previous studies of reading interventions for older students suggest that larger group sizes are just as effective as smaller groups for this population (Vaughn et al., 2010). During both

school years, each school had a different bell schedule, but students received approximately 3.75 to 4.25 hours of intervention weekly. Students continued to receive their core classes (i.e., English, math, science, social studies) and attended the RIA in place of an elective class.

Phase I. RIA was delivered in two phases. During the first semester of year 1, students participated in Phase I of the intervention. This phase addressed word study, fluency, and vocabulary through REWARDS Secondary, an explicit instruction program designed for students in grades 6 through 12 (Archer, Gleason, & Vachon, 2014). Reading interventionists implemented REWARDS Secondary 3 to 4 days each week. The program taught students a flexible strategy for decoding multisyllabic words that involves (a) looking for prefixes, suffixes, and vowels; (b) saying the parts of a word; (c) saying the whole word; and (d) making it a real word. Students developed automaticity by reading these words in isolation and in connected text. Vocabulary knowledge was promoted through teaching affixes and academic vocabulary related to students' content area instruction.

In the same semester reading interventionists also introduced the procedures for Collaborative Strategic Reading (CSR; Klingner & Vaughn, 1998) and explicitly taught mini-lessons for each of the CSR components (i.e., preview, click and clunk, get the gist, questioning, review). The mini-lessons were taught one or two times each week before or after the REWARDS Secondary lessons. Students then learned to work collaboratively in small groups of approximately four students to use these strategies with content area text. Students were given a CSR flip chart with the following resources: fix-up strategies; prefix, suffix, and root word lists; a Spanish-English cognate list; and question starters.

Phase II. Phase II of the intervention, which focused on vocabulary and comprehension, began in the second semester of year 1 and continued through all of year 2. Through 14

instructional units corresponding with content area instruction in science and social studies, Phase II emphasized the application of practices previously taught for content area text. Each unit activated and built on prior knowledge through CSR Preview, where students brainstormed about the topic and viewed short video clips. Students then read the headings and subheadings, completed an anticipation-reaction guide graphic organizer, and recorded the purpose for reading in their CSR learning log. Six to eight new vocabulary words were explicitly taught using a six-step procedure: (a) have students pronounce the word, (b) provide a student-friendly definition with visual, (c) have students discuss what is known about the word, (d) provide examples and nonexamples, (e) engage in deep-processing activities, and (f) have students create powerful sentences with the new word (Vaughn Gross Center for Reading and Language Arts at The University of Texas at Austin, 2010). Vocabulary words were reviewed each day throughout the unit and students completed graphic organizers (e.g., word maps), identified word relatives, and participated in structured question prompts that used the words. Students then worked collaboratively in their small groups to identify unknown words and use fix-up strategies to determine their meanings, generate main ideas, ask and answer questions, and create review statements for the whole text. After reading, students wrote summaries using a strategy adapted from the macrorules for summarizing text (Brown & Day, 1983). Reading interventionists then used a graphic organizer and discussion protocol guide to lead a formal discussion with students. Last, students had free-choice reading time, in which they selected high-interest books and magazines to read for 10 minutes each class period. Throughout Phase II, students continued to practice reading multisyllabic words and reviewed meanings of common affixes learned in Phase I.

Reading Interventionists and Training

The research team hired and trained five reading interventionists to deliver RIA. All of them were female, had state certifications in secondary reading or English language arts, and had a minimum of 5 years of teaching experience. Each year, interventionists completed 40 hours of preintervention training, which focused on the elements of effective instruction (i.e., explicit instruction), as well as the implementation of REWARDS Secondary and CSR (Archer et al., 2014; Klingner & Vaughn, 1998). Depending on the need, an additional 8 to 16 hours of training was provided at the end of Phase I to prepare teachers to implement the content area units in Phase II. Two members of the research team provided limited coaching and support throughout the year in the form of in-person and audio coaching observations. The interventionists and research team members held biweekly phone conferences to discuss student progress and adjust instruction as necessary.

Implementation Fidelity

Fidelity data were collected in three separate 2-week waves during each year of intervention for six total fidelity collection periods. The reading interventionists recorded two of their reading intervention class periods on a digital audio recorder and then uploaded the files to a secure server. An implementation fidelity protocol was used to determine the adherence to and quality of each recorded lesson. Intervention adherence was rated on a Likert-type scale ranging from 0 (*not observed*) to 4 (*high*) for each of the intervention components (i.e., word study, vocabulary, comprehension, discussion/interpretation of text, motivation). Quality of overall reading intervention, RIA, and classroom management were also rated on a Likert-type scale ranging from 1 (*lowest quality*) to 5 (*highest quality*). Four raters were trained on the fidelity coding procedures, and the project coordinator served as the "gold standard" (Gwet, 2001). Each rater had to score 90% or higher reliability with the gold standard before coding independently.

Forty percent of each interventionist's audio files were coded. Mean fidelity scores for each year are reported in Table 4. Average intervention adherence ranged (on a 4-point scale) from 2.48 to 2.97 for year 1 and 2.93 to 3.71 for year 2. Intervention adherence ratings were comparable to ratings from the previous study of RIA with monolingual students (Vaughn, Roberts, Wexler, et al., 2015) in which the range (on a 5-point scale) was 2.8 to 3.4.

Considering that the fidelity data revealed less than ideal implementation of the program, though in line with a previous study with monolingual high school students, questions regarding why the teachers were not "better coached" to improve fidelity are appropriate. We consistently monitored teachers' fidelity and provided feedback to teachers including modeling lessons and guiding their implementation with fidelity improving over time. The challenges for these teachers were related to the "reality" of teaching high school students from extremely high poverty homes with extraordinary challenges including: (a) multiple family members incarcerated, missing, and reappearing in their lives, (b) inadequate resources related to low-income including services turned off in their homes, having to move quickly to avoid bills, etc., and (c) personal stresses that are common among high school students from families who are struggling with the daily challenges of survival. We do not exaggerate these problems and our teachers were aware that they were teaching youngsters that required them to adjust the instruction or on some days abandon it completely to respond to the personal stresses of their students.

Comparison Condition

Students in the comparison condition were enrolled in a variety of different courses, while treatment students received the RIA. Some participants received an extension of the school-provided English I or II course, in which the teacher continued instruction from the

previous lesson. Other students were enrolled in electives such as Introduction to Cosmetology, Principles of Information Technology, Concepts of Engineering and Technology, Art, Welding, JROTC, etc. After a thorough investigation of participants' course schedules and discussions with the school administrators, we have no evidence to suggest that the comparison condition received a comparable reading intervention and it is appropriate to refer to the condition as BAU.

Data Collection and Measures

Participants were provided reading assessments (i.e., word reading, vocabulary, and comprehension) in September and May, at the beginning and end of each year of the intervention. The research team hired and trained test administrators to assess the participants, and each test administrator received approximately 20 hours of training on test administration and data collection procedures. Before administering assessments, the test administrators had to have an interrater reliability above .90 for each subtest administered.

Distal Measures

Test of Word Reading Efficiency–Second Edition. Both subtests of the TOWRE-2 (Torgesen et al., 2012) were administered individually to all participants. The first subtest, the sight word subtest, is a timed assessment that measures the number of real printed words that a student can read in 45 seconds. The second subtest, the phonemic decoding subtest, assesses a student's ability to accurately read a list of nonwords in 45 seconds. Test-retest reliability ranges from .83 to .96 and alternative-form reliability exceeds .90 for both subtests.

Gates-MacGinitie reading comprehension and vocabulary subtests. Reading comprehension was assessed through the comprehension subtest of the Gates-MacGinitie Reading Tests (4th ed.; GMRT-4), which assesses a student's ability to answer literal and inferential multiple-choice questions from 11 narrative and expository passages (MacGinitie,

MacGinitie, Maria, Dreyer, & Hughes, 2000). It is a timed and group-administered test. Form S was given at pretest and Form T was given at posttest. Alternate-form reliability is 0.80 and Kuder-Richardson Formula 20 reliability is .93 for the fall administration of Form S and .92 for the spring administration of Form T (MacGinitie et al., 2000).

The GMRT-4 vocabulary subtest is a standardized, timed measure that includes 45 vocabulary questions based on reading passages. Form S was administered at pretest and Form T administered at posttest. The alternate-form reliability is. 83 and Kuder-Richardson Formula 20 reliability is .92 for the fall administration of Form S and .91 for the spring administration of Form T.

Test of Silent Reading Efficiency and Comprehension. Comprehension was also assessed through the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010), in which students read as many sentences as possible in a 3-minute period and determine the truthfulness of each sentence. Form A was administered at pretest in the fall and Form C was administered at posttest in the spring. Alternate-form reliability exceeds .85.

Proximal Measure

RIA vocabulary. The second vocabulary assessment was a researcher-created proximal vocabulary measure that assessed students' knowledge of words taught in the intervention. Academic vocabulary words (e.g., *determine*, *progressive*) were selected from the social studies and science content area units in Phase II of the intervention. The assessment consists of 29 words divided into four sections, and students were instructed to match the words with their definitions.

Plan for Analysis

We used multigroup latent growth analysis (MG-LGA) to evaluate the effects of treatment (Muthén & Curran, 1997). MG-LGA estimates effects as the difference in change experienced by two or more groups, treatment and BaU in this case. Generally, growth models yield an intercept, which is often an estimate of status at Time 1, and a slope, which describes change. Variances for intercept and slope, covariances between intercept and slope, and error terms are estimated as well. In the multigroup case, data from the two groups are analyzed separately though simultaneously. A freely estimated model results in two unique sets of the usual complement of model parameters (i.e., intercept and slope for each group, variances, etc.). A constrained model requires that all parameters or a subset of the two sets of parameters be estimated as equal across conditions. In MG-LGA, the intercept, slope, parameter variances, and covariances for treatment and BaU are constrained as equal (see Figure 1), and a third growth factor unique to the treatment accounts for change in the scores over time in addition to the growth trajectory of the BaU group. In an RCT, this third parameter represents the effect of treatment. We also modeled the relationship of intercept and change over time and the extent to which the relationship may have differed across conditions. Some treatments work for all targeted participants. Some work for only a subset of students; these subsets can be identified using statistical moderation (Fairchild & MacKinnon, 2009). In this case, we were primarily interested in the effect of pretreatment status on change and the treatment-related differences in that relationship. We considered the significance of moderated effects at the mean for each moderator and at values that correspond to 2, 1.5, 1, and .5 standard deviations above and below the sample (or group) mean. We also examined the comparability of groups at Time 1 by relaxing the equality constraints on the groups' intercepts and comparing the relative fit (in terms of $\Delta\chi^2$) of nested models (Bovaird, 2007). The fit of individual models was evaluated using the

comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) following recommendations of Hu and Bentler (1999). We treated missing data as missing at random and used a full information maximum likelihood estimator to fit models. Effect sizes (Hedges' *g*) were estimated as the difference in slope coefficients across the two groups divided by the pooled standard deviation at Time 1 (Feingold, 2009).

Results

Means and standard deviations for reading outcomes are summarized in Table 5. Differing sample sizes across measurement occasions reflect attrition. The variables were normally distributed based on estimates of skewness and kurtosis.

Growth Trends in Reading

We addressed the research questions by building the best fitting full model for each outcome under the assumption that a well fit two-group model with the indicated constraints was a reasonable baseline against which to compare alternative models (Curran & Muthén, 1999). The first step was to establish attributes of normative change by fitting models in BaU. The traditional linear model, where the slope parameter predicts equally spaced or time-aligned integer values (e.g., 0, 1, 2, 3…k), did not track with data observed in BaU, due partly to the large decrease in average scores from the end of 9th grade to the beginning of 10th grade. The loss of reading skill among lower socioeconomic status groups during the summer months is well documented (Alexander, Entwisle, & Olson, 2007; Allington & McGill-Franzen, 2003; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Downey, Von Hippel, & Broh, 2004). When modeling more than 2 years of literacy data (i.e., multiple summers) using more than four time points, polynomial models (e.g., quadratic form) will sometimes accommodate the average deceleration represented by loss over multiple summers. Alternating periods of treatment and

nontreatment (i.e., school months and summer months) can also be indicated with time-varying covariates (McCoach & Kaniskan, 2010), models that estimate the average trend(s) during period(s) of active treatment, during the summer months, and across the entire time span. Piecewise growth models represent a third analytic alternative.

However, summer learning loss was not our interest, per se. Our purpose, instead, was to represent change in BaU with the greatest economy for the purpose of "removing" or controlling expected change in the treatment group, thereby allowing estimates of the treatment's unique effect. We used a combination of fixed and freed time scores for this purpose (Muthén & Curran, 1997; Curran & Muthén, 1999; Wu, Liu, Gadermann, & Zumbo, 2010). Models for all outcomes were fit by allowing scores for Time 2 (end of 9th grade) and Time 3 (beginning of 10th grade) to freely estimate in BaU. Fixing Time 1 (beginning of 9th grade) as 0 and Time 4 (end of 10th grade) as 1 per a standard linear growth model (where Time 2 would be fixed at .25 and Time 3 fixed at .75), scores for the two other measurement occasions were freely estimated based on the observed data. These predicted values represent time points, but they can be expressed as estimates for average performance at the time point(s) of interest (for intercept growth factor equal to 0, the predicted outcome for time *t* is the product of the model-estimated time score and the mean for the slope growth factor). Note that slope in these models describes the average rate of change for a time score change of 1, which differs from its meaning in a classically linear model, where the growth factor is assumed constant across all time points. However, slope in both cases (in the linear model and in models with freely estimated time scores) represents a meaningful and efficient index of average change in BaU over the 2-year time period. Further, to the extent that the multigroup model with the BaU-estimated time scores imposed on the treatment group (i.e., the constrained model) fits the observed data, the slope parameter describes

expected change in the treatment group, thereby isolating variance that is unique to cases allocated to intervention. Model fits for the multigroup, constrained models are summarized in Table 6 for each outcome. To evaluate the integrity of randomization, we relaxed constraints on groups' equivalence at Time 1 (i.e., intercept as equal across BaU and treatment). At pretest, the groups differed on the TOWRE-2 sight word subtest ($\Delta\chi^2 = 6.1$ (1), $p = .01$), the GMRT-4 comprehension subtest ($\Delta\chi^2 = 11.2$ (1), $p = .00$), and the TOWRE phonemic decoding subtest ($\Delta\chi^2 = 5.6$ (1), $p = .02$). We included these differences in the final models reported in Table 6.

Treatment Effects

For the TOSREC, the normative slope was 2.31 ($p < .01$), which represents average change in scale score units per unit of *t*, or change that is independent of intervention. The treatment slope, 2.52 ($p < .05$), describes change that is unique to the treated group. We argue that these estimates have high internal validity given the excellent fit for the multigroup model (χ^2 (9, 353) = 4.98, $p = .84$; CFI = 1.00; TLI = 1.00; RMSEA = .00 [90% CI = .00, .05]) and the randomized design that underlie these data. The effect size (Hedges' *g*) for the TOSREC was .18.

For the proximal RIA vocabulary measure, the multigroup model fit the data well (χ^2 (9, 353) = 10.93, $p = .28$; CFI = .99; TLI = .99; RMSEA = .04 [90% CI = .00, .09]). The normative slope was 1.36 ($p < .001$). Slope in the treatment group (beyond the normative slope) was 1.00 ($p < .01$). The effect size was .41.

The model fit for the GMRT-4 reading comprehension subtest was marginal (χ^2 (8, 353) = 19.39, $p = .201$; CFI = .93; TLI = .90; RMSEA = .09 [90% CI = .04, .14]). The normative slope was 32.96 (< .001), expressed in extended scale score points. The treatment-specific mean slope did not differ from 0 (-2.49; $p = .47$). The effect size was -.09. Results for the GMRT-4

vocabulary subtest were similar (effect size = .00), although the model was a better fit for the data (χ^2 (9, 353) = 13.98, p = .13; CFI = .98; TLI = .97; RMSEA = .06 [90% CI = .00, .11]).

The models for the two TOWRE-2 subtests were very good. For sight word reading (χ^2 (8, 353) = 3.37, p = .91; CFI = 1.00; TLI = 1.00; RMSEA = .00 [90% CI = .00, .04), the normative change was 2.80 (p < .001) standard score points. The treatment slope was -0.22 (p = .84). Note, however, that the two groups were not comparable at Time 1, on average. Initial status for the treatment group was just over 3 standard score points lower (-3.04, p < .01) than the BaU condition. The effect size for sight word reading was -.02, and the effect size for phonemic decoding was .06.

Moderating Effects

We regressed treatment-group slopes on their respective intercepts for each outcome. We evaluated the extent to which each coefficient differed from 0 using the BaU-specific coefficient (fixed at 0) as the comparison to identify subgroups for which the intervention may have been particularly beneficial and the outcomes on which they may have excelled. We focus on those outcomes with well fit multigroup models (i.e., TOSREC, RIA vocabulary, GMRT-4 vocabulary subtest, TOWRE-2 sight word and phonemic decoding subtests). The interaction of slope on intercept for the RIA vocabulary measure was statistically significant from 0 at its mean (p = .004) and for all Time 1 scores at or above the sample mean, suggesting that treated students with higher pretest scores (scores at or above the sample average) on the RIA measure made greater gains on the RIA over time than did similar students in BaU. The opposite pattern was evident with the TOSREC—low-scoring students in treatment outperformed higher scoring students in treatment when compared to lower and higher scoring students in BaU.

Discussion

This study examined the efficacy of a 2-year extensive reading intervention, RIA, on the reading achievement of ELs identified as struggling readers in high school based on their performance on the reading portion of the eighth-grade state accountability test. Using an RCT, we determined that students assigned to the reading treatment for 2 years demonstrated unique treatment effects on the TOSREC, a timed sentence reading measure that assesses fluency and comprehension. Additionally, students in the treatment condition realized treatment effects as measured by the RIA vocabulary assessment. There were no significant treatment effects on the passage reading comprehension measure or the word reading measure.

Considering the extensive treatment over 2 years, it is reasonable to expect that students in the treatment condition would perform differentially better on the reading comprehension measure than those in the BaU condition. There are several considerations for interpreting this null finding. First, students in this sample overwhelmingly spoke a language other than English in the home (99.5%). Though this could potentially enhance their opportunities in the future with respect to maintaining a language in addition to English, these students had underdeveloped vocabulary in English that was difficult to compensate for, even with extensive instruction. According to the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), reading comprehension is the product of decoding (i.e., word reading) and linguistic development, which is largely determined through vocabulary development. As texts becomes more complex, the role of vocabulary development increases in importance such that students with underdeveloped vocabulary are increasingly challenged to understand and learn from text. Though vocabulary instruction is associated with improved outcomes on the vocabulary taught, it is rarely associated with broader outcomes related to vocabulary or comprehension development (August, Carlo, Dressler, & Snow, 2005; Vaughn, Martinez et al., 2009). Based on

this finding, we believe that it is critical to enhance vocabulary development as a means of improving comprehension but that a 50-minute reading intervention is unlikely to be sufficiently powerful. As suggested by others, enhanced vocabulary knowledge requires a schoolwide approach, such as word generation (Lawrence, Crosson, Pare-Blagoev, & Snow, 2015; Snow, Lawrence, & White, 2009), that promotes vocabulary and language development across the curriculum and is implemented throughout schooling.

Second, despite the lack of differential treatment effects on passage comprehension, it is relevant to consider that students in both conditions, treatment and BaU, made considerable growth from pretest (beginning of 9th grade) to posttest (end of 10th grade) on passage comprehension. Both groups made nearly an 8-point standard score gain, moving from more than 1 to .75 standard deviations below the normative sample. Thus, although this gain was not differentially significant, it is practically significant in that students may be reaching a threshold of reading comprehension that allows them to benefit from text reading previously not available to them. Lemons et al. (2014) offer another explanation for the enhanced performance of the BaU students that may relate to this study. They argue that many contextual factors, including priorities to improve reading performance of students (e.g., accountability measures), have established a school context that influences the quantity and quality of reading instruction provided to the BaU students. It is noteworthy that the schools participating in this study realized that their students demonstrated low performance in reading and provided enhanced reading opportunities to many students in the BaU condition. This explanation for enhanced BaU performance has been offered as a possible explanation for why effect sizes have declined in experimental reading interventions over time (Scammacca et al., 2016).

Whereas analyses revealed no differential treatment effects on passage comprehension, there were unique effects for students in the treatment condition on sentence reading fluency and comprehension. The burden of underdeveloped language on successful understanding is smaller for sentence reading than it is for passage reading, which may help explain why students demonstrated differential performance on the sentence reading task. The sentence reading measure also best illustrates the summer loss in both the treatment and BaU groups, presenting the very real challenge of improving outcomes over time with the considerable summer loss experienced by all participants. Though summer loss in high-poverty groups has been previously documented (Alexander et al., 2007; Allington & McGill-Franzen, 2003; Cooper et al., 1996; Downey et al., 2004), we consider the summer loss measured by the TOSREC (Wagner et al., 2010) a noteworthy finding that warrants consideration in future studies. The treatment students made standard score gains of 6 points between fall and spring of their ninth-grade year (SS = 75 to SS = 81) and then lost 7 points over the summer (SS = 74). Again, the treatment students realized on average a 5-point SS gain during the 10th grade year (SS = 79). Assuming this "yo yo" of improving quite strongly during the school year and then regressing so significantly in the summer is representative of students' reading progress over time, then this summer loss warrants serious consideration for what beneficial supports can be provided to maintain or enhance summer reading performance.

This research study also examined the moderating effects of treatment to determine whether there were subgroups of students for whom the intervention was distinctly beneficial. Students with relatively higher RIA vocabulary scores at pretest made greater gains than students in the BaU condition, whereas students with relatively lower TOSREC (Wagner et al., 2010) scores at pretest outperformed students with relatively higher scores. The TOSREC may be a

measure of reading mastery as students with relatively lower TOSREC scores at pretest outperformed students in the BAU condition with relatively higher scores. The opposite effect was observed for the vocabulary measure in which students who started with higher vocabulary scores gained the most. It may be that vocabulary acquisition occurs at a more rapid rate when you already know a lot of words than when you have a more minimal English vocabulary.

We position our findings within the context of a synthesis reported by Herrera, Truckenmiller, and Foorman (2016) who report on 20 years of research on reading interventions for adolescents. Over this time period, they identified only 12 studies as having positive or potentially positive outcomes and 9 of these 12 studies were implemented during English language arts or content area instruction – not as interventions for students with significant reading difficulties. Furthermore, they reported that they identified 0 studies at the high school level that met their criteria.

Limitations

A limitation of this study is the high attrition rate. As described previously in this paper, we needed to identify the sample of students for the study during their eighth-grade year, the academic year prior to the study, to be able to randomize the students and allow the high schools to assign class schedules. Because students were permitted to attend the high school of their choosing, about one-half of the randomized sample did *not* enroll in the fall at one of the participating high schools that represented their feeder pattern. As we expected overall high attrition rates from eighth to ninth grade, we oversampled accordingly, and the study is adequately powered to detect treatment effects on reading achievement outcomes that are typical for 2-year reading interventions with older struggling readers. Nonetheless, the high rate of overall attrition represents a threat to the findings' internal validity.

Another limitation of the study is the low fidelity of implementation. Conducting RCTs in high school settings is extraordinarily challenging and perhaps explains why recent meta-analyses find few or no RCTs conducted with high school students (Scammacca et al., 2016). These interventions were implemented in inner-city schools, where the majority of families have very few resources, discipline problems are schoolwide, and teacher turnover is high. Many of the participating students had responsibilities outside of school, including jobs, and they slept or tried to sleep during classes. All of these factors contributed to challenges implementing the intervention with fidelity.

Implications for Future Research

This study contributes to the research on secondary school ELs with reading difficulties and identifies many of the challenges to supporting ELs in acquiring proficient reading comprehension, including expanding language and vocabulary acquisition while simultaneously improving text reading. The significant summer loss experiences by students also points to a direction for future research. As we reflect on our experiences and the data from the students in these inner-city schools, we are hard-pressed to think of any single intervention that could possibly remedy the significant language and literacy challenges these students experience. Instead, we conclude that future directions for research require extensive schoolwide efforts focused on improving language and literacy across the curriculum over many years, with feasible approaches to addressing summer loss that take into account the economic challenges of the target group (i.e., the likelihood that they need to work in the summer).

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Summer learning and its implications: Insights from the Beginning School Study. In R. Fairchild & G. Noam (Eds.), *Summertime: Confronting risks, exploring solutions* (Vol. 114, pp. 11–32). San Francisco, CA: Jossey-Bass.

Allington, R. L., & McGill-Franzen, A. M. (2003). The impact of summer loss on the reading achievement gap. *Phi Delta Kappan, 85*(1), 68–75.

Archer, A. L., Gleason, M. M., & Vachon, V. L. (2014). *REWARDS secondary* (3rd ed.). Longmont, CO: Cambium Learning Group Sopris Learning.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*(1), 50-57.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/PracticeGuide/19

Berliner, B., Barrat, V. X., Fong, A. B., & Shirk, P. B. (2008). *Reenrollment of high school dropouts in a larger, urban school district* (Issues & Answers Report, REL 2008-No. 056). Washington, DC: U.S Department of Education, Institute of Educational Sciences,

National Center for Education Evaluation and Regional Assistance, Regional Educational Laboaratory West. Retrieved from http://ies.ed.gov/ncee/edlabs

Boardman, A. G., Klingner, J. K., Buckley, P., Annamma, S., & Lasser, C. J. (2015). The efficacy of Collaborative Strategic Reading in middle school science and social studies classes. *Reading and Writing*, *28*(9), 1257–1283.

Boardman, A. G., Vaughn, S., Buckley, P., Reutebuch, C., Roberts, G., & Klingner, J. (2016). Collaborative Strategic Reading for students with learning disabilities in upper elementary classrooms. *Exceptional Children*, *82*(4), 409–427.

Boulay, B., Goodson, B., Frye, M., Blocklin, M., & Price, C. (2015). *Summary of research generated by Striving Readers on the effectiveness of interventions for struggling adolescent readers* (NCEE 2016-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In T. Little, J. Bovaird, & N. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 149–182). Mahwah, NJ: Erlbaum.

Brasseur-Hock, I. F., Hock, M. F., Kieffer, M. J., Biancarosa, G., & Deshler, D. D. (2011). Adolescent struggling readers in urban schools: Results of a latent class analysis. *Learning and Individual Differences, 21*(4), 438-452. doi: 10.1016/j.lindif.2011.01.008

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1–14.

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing, 26*(7), 1059-1086. doi: 10.1007/s11145-012-9406-3

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227–268.

Corrin, W., Somers, M. A., Kemple, J. J., Nelson, E., & Sepanik, S. (2008). *The enhanced reading opportunities study: Findings from the second year of implementation* (NCEE 2009-4036). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Curran, P. J., & Muthén, B. O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology, 27*(4), 567–595.

Downey, D., Von Hippel, P., & Broh, B. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*(5), 613–635.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. T., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research, 79*(1)262-300. doi: 10.3102/0034654308325998

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderating effects. *Prevention Science*, *10*(2), 87–99.

Feingold, A. (2009*)*. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14,* 43–53.

Flynn, L. J., Zheng, X., & Swanson, H. L. (2012). Instructing struggling older readers: A selective meta-analysis of intervention research. *Learning Disabilities Research & Practice*, *27*, 21–32. doi:10.1111/j.1540-5826.2011.00347.x

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Retrieved from http://www.centeroninstruction.org

Genesee, F., Lindholm-Leary, K., Saunders, W. M., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. New York, NY: Cambridge University Press.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10.

Gwet, K. (2001). *Handbook of inter-rater reliability*. Gaithersburg, MD: STATAXIS.

Hall, C., Roberts, G. J., Cho, E., McCulley, L. V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review*, *29*(4), 763–794.

Hernandez, D. J. (2011). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. Retrieved from http://www.aecf.org/resources/double-jeopardy/

Herrera, S., Truckenmiller, A. J., and Foorman, B. R. (2016). *Summary of 20 years of research on the effectiveness of adolescent literacy programs and practices* (REL 2016–178). Washing ton, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Hoover, W. and Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal,* 2, 127–160.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

James-Burdumy, S., Deke, J., Gersten, R., Lugo-Gil, J., Newman-Gonchar, R., Dimino, J., . . . Liu, A. Y. H. (2012). Effectiveness of four supplemental reading comprehension interventions. *Journal of Research on Educational Effectiveness*, *5*, 345–383. doi:10.1080/19345747.2012.698374

Kemple, J. J., Corrin, W., Nelson, E., & Salinger, T. (2008). *The enhanced reading opportunities study: Early impact and implementation findings* (NCEE 2008-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Kennelly, L., & Monrad, M. (2007). Approaches to dropout prevention: Heeding early warning signs with appropriate interventions. Washington, DC: National High School Center at the American Institutes for Research.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*(4), 851–868. doi:10.1037/0022-0663.100.4.851

Kieffer, M. J. (2011). Converging trajectories: Reading growth in language minority learners and their classmates, kindergarten to grade eight. *American Educational Research Journal, 48*, 1157–1186.

Klingner, J. K., & Vaughn, S. (1998). Using Collaborative Strategic Reading. *Teaching Exceptional Children, 30*(6), 32–37.

Kundert, D. K., Newman, D. L., Gifford, T. A., Haase, R. F., & Clure, G. (2012). *New York State Striving Readers review technical report 3: The impact of REWARDS on multisyllabic word reading.* Albany, NY: University at Albany/SUNY: The Evaluation Consortium.

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., Snow, C. (2015). Word generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, *52*(4), 750–786

Lesaux, N. K., & Harris, J. R. (2017). An investigation of comprehension processes among adolescent English learners with reading difficulties. *Topics in Language Disorders*, *37*(2), 182–203.

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal*, *47*(3), 596–632.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal, 51*(6), 1159–1194. doi:10.3102/0002831214532165

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: Riverside.

Mancilla-Martinez, J., Kieffer, M. J., Biancarosa, G., Christodoulou, J. A., & Snow, C. E. (2011). Investigating English reading comprehension growth in adolescent language minority learners: Some insights from the simple view. *Reading and Writing, 24*(3), 339–354. doi:10.1007/s11145-009-9215-5

McCoach, D. B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in Psychology, 1*, 1–12.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2*(4), 371–402.

National Center for Education Statistics [NCES]. (2016). *2015 reading assessment.* Retrieved from the NAEP Data Explorer from http://nces.ed.gov/nationsreportcard/naepdata

National Research Council and National Academy of Education. (2011). High School dropout, graduation, and completion rates: Better data, better measures, better decisions. Committee for Improved Measurement of High School Dropout and Completion Rates: Expert Guidance on Next Steps for Research and Policy Workshop. R.M. Hauser and J.A. Koenig, (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Roberts, G., Vaughn, S., Fletcher, J., Stuebing, K., & Barth, A. (2013). Effects of a response-based, tiered framework for intervening with struggling readers in middle school. *Reading Research Quarterly*, *48*(3), 237–254.

Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4–

12, 1914–2014. *Review of Educational Research, 86*(3), 756-800. doi: 10.3102/0034654316652942

Shippen, M. E., Houchins, D. E., Steventon, C., & Sartor, D. (2005). A comparison of two direct instruction reading programs for urban middle school students. *Remedial and Special Education*, *26*(3), 175–182.

Short, D., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners: A report to Carnegie Corporation of New York*. New York, NY: Alliance for Excellent Education.

Snow, C. E., Lawrence, J. F., White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2,* 325–344.

Snyder, T.D., de Brey, C., and Dillow, S.A. (2016). *Digest of education statistics 2015.* (NCES 2016-014). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Texas Education Agency. *2015-2016 school report card.* Retrieved from http://www.tea.texas.gov

Torgesen, J. K., Wagner R. K., & Rashotte C. A. (2012). Test of word reading efficiency–Second edition. Austin, TX: Pro-Ed.

University of Minnesota. (2015). *Check & connect*. Retrieved from http://www.ici.umn.edu/checkandconnect

Vaughn Gross Center for Reading and Language Arts at The University of Texas at Austin. (2010). *Vocabulary and oral language development*. Austin, TX: Author.

Vaughn, S., Chard, D. J., Bryant, D. P., Coleman, M., Tyler, B.-J., Linan-Thompson, S., & Kouzekanani, K. (2000). Fluency and comprehension interventions for third-grade students. *Remedial and Special Education, 21*(6), 325–335.

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., & Stillman-Spisak, S. J. (2011). Efficacy of Collaborative Strategic Reading with middle school students. *American Educational Research Journal*, *48*(4), 938–964.

Vaughn, S., Martinez, L. R., Linan-Thompson, S. L., Reutebuch, C. K, Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness, 2*, 297-324. doi: 10.1080/19345740903167018

Vaughn, S., Roberts, G., Schnakenberg, J. B., Fall, A., Vaughn, M. G., & Wexler, J. (2015). Improving reading comprehension for high school students with disabilities. *Exceptional Children, 82*(1), 117–131. doi:10.1177/0014402915585478

Vaughn, S., Roberts, G., Wexler, J., Vaughn, M. G., Fall, A., & Schnakenberg, J. B. (2015). High school students with reading comprehension difficulties. *Journal of Learning Disabilities, 48*(5), 546–558. doi:10.1177/0022219413515511

Vaughn, S., Wanzek, J., Wexler, J., Barth, A., Cirino, P. T., Fletcher, J., . . . Francis, D. (2010). The relative effects of group size on reading progress of older students with reading difficulties. *Reading & Writing, 23*(1), 931–956. doi:10.1007/s11145-009-9183-9

Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., Barth, A., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth-grade students with persistently

inadequate response to intervention. *Journal of Learning Disabilities, 45*(6), 515-525. doi:10.1177/0022219411402692

Vaughn, S., Wexler, J., Roberts, G., Barth, A., Cirino, P., Romain, M., . . . Denton, C. (2011). Effects of individualized and standardized interventions on middle school students with reading disabilities. *Exceptional Children*, *77*, 391–407.

Wagner, R. K., Torgesen, J., Rashotte, C. A., & Pearson, N. (2010). *Test of sentence reading efficiency and comprehension.* Austin, TX: Pro-ED.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541–561.

Wanzek, J., Vaughn, S., Roberts, G., & Fletcher, J. M. (2011). Efficacy of a reading intervention for middle school students with learning disabilities. *Exceptional Children, 78*(1), 73–87. doi:10.1177/001440291107800105

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review*, *28*(3), 551–576.

Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murray, C. S., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*(2), 163-195. doi: 10.3102/0034654313477212

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: a synthesis of 20 years of research. *Reading and Writing, 23*(8), 889-912. doi: 10.1007/s11145-009-9179-5

Wexler, J., Vaughn, S., Roberts, G., & Denton, C. A. (2010). The efficacy of repeated reading and wide reading practice for high school students with severe reading disabilities. *Learning Disabilities Research and Practice*, *25*(1), 2–10. doi:10.1111/j.1540-5826.2009.00296.x

What Works Clearinghouse (2017). Procedures Handbook, Version 4.0, 7/14/2017.

Wu, A. D., Liu, Y., Gadermann, A. M., & Zumbo, B. D. (2010). Multiple-indicator multilevel growth model: A solution to multiple methodological challenges in longitudinal studies. *Social Indicators Research: International Interdisciplinary Journal for Quality of Life Measurement, 97,* 123–142.

Table 1

Student Demographics

	Overall (n = 358)		Treatment (n = 175)		Comparison (n = 183)	
	n	%	*n*	%	*n*	%
Race						
African American	3	.8	3	1.7	0	0
Hispanic	320	89.4	154	88	166	90.7
American Indian	1	.3	0	0	1	.5
White	1	.3	0	0	1	.5
Missing	33	9.2	18	10.3	15	8.2
Gender						
Male	188	52.5	89	50.9	99	54.1
Female	137	38.3	68	38.9	69	37.7
Missing	33	9.2	18	10.3	15	8.2
Home language						
Spanish	320	89.4	153	87.4	167	91.3
English	1	.3	1	.6	0	0
Other	4	1.1	3	1.7	1	.5
Missing	33	9.2	18	10.3	15	8.2
English learner status						
Former	132	36.9	59	33.7	73	39.9
Current	226	63.1	116	66.3	110	60.1
Economically disadvantaged						
No	55	15.4	28	16	27	14.8
Yes	270	75.4	129	73.7	141	77
Missing	33	9.2	18	10.3	15	8.2
Special education services						
No	281	78.5	136	77.7	145	79.2
Yes	44	12.3	21	12.0	23	12.6
Missing	33	9.2	18	10.3	15	8.2

Table 2
Overall and Differential Attrition Rates Across the Course of the Study

	Time 1	Time 2	Time 3	Time 4
	%	%	%	%
TOSREC				
Overall attrition	51.30	56.40	58.1	63.4
Differential attrition	-2.7	-1.5	-2.7	-5.4
Gates comprehension				
Overall attrition	51.30	55.40	57.1	63
Differential attrition	-3.2	-1.2	-2.4	-4.6
Gates vocabulary				
Overall attrition	51.30	55.40	57.1	63
Differential attrition	-2.7	-1.2	-2.4	-4.6
TOWRE-2 sight word				
Overall attrition	51.30	55.50	57.1	63
Differential attrition	-2.7	-0.9	-2.4	-4.6
TOWRE-2 phonemic decoding				
Overall attrition	51.50	55.40	57.1	63
Differential attrition	-2.4	-1.2	-2.4	-4.6
RIA vocabulary				
Overall attrition	51.2	55.4	57.1	63
Differential attrition	-3	-1.2	-2.4	-4.6

Note. TOSREC = Test of Silent Reading Efficiency and Comprehension; Gates = Gates-MacGinitie Reading Tests; TOWRE-2 = Test of Word Reading Efficiency–Second Edition; RIA = Reading Intervention for Adolescents.

Table 3

Effect Size Estimates for the Baseline Equivalence

		95% CI	
	Effect size	Lower	Upper
TOSREC	0.09	-1.36	1.55
Gates comprehension	-0.18	-2.95	2.59
Gates vocabulary	-0.01	-2.46	2.43
TOWRE-2 sight word	0.26	-0.93	1.44
TOWRE-2 phonemic decoding	0.19	-1.12	1.50
RIA vocabulary	-0.14	-0.40	0.12
STAAR	0.03	-5.18	5.24

Note. CI = confidence interval; TOSREC = Test of Silent Reading Efficiency and Comprehension; Gates = Gates-MacGinitie Reading Tests; TOWRE-2 = Test of Word Reading Efficiency–Second Edition; RIA = Reading Intervention for Adolescents; STAAR = State of Texas Assessments of Academic Readiness.

Table 4

Mean Fidelity Scores for Observed Variables by Year

Observed component	Year 1	Year 2
Word study adherence	2.97	3.71
Vocabulary adherence	2.74	3.43
Comprehension adherence	2.77	3.32
Discussion interpretation of text adherence	2.68	2.93
Motivation engagement adherence	2.48	3.17

Note. Scores rated on a 4-point scale, with 4 as the highest score.

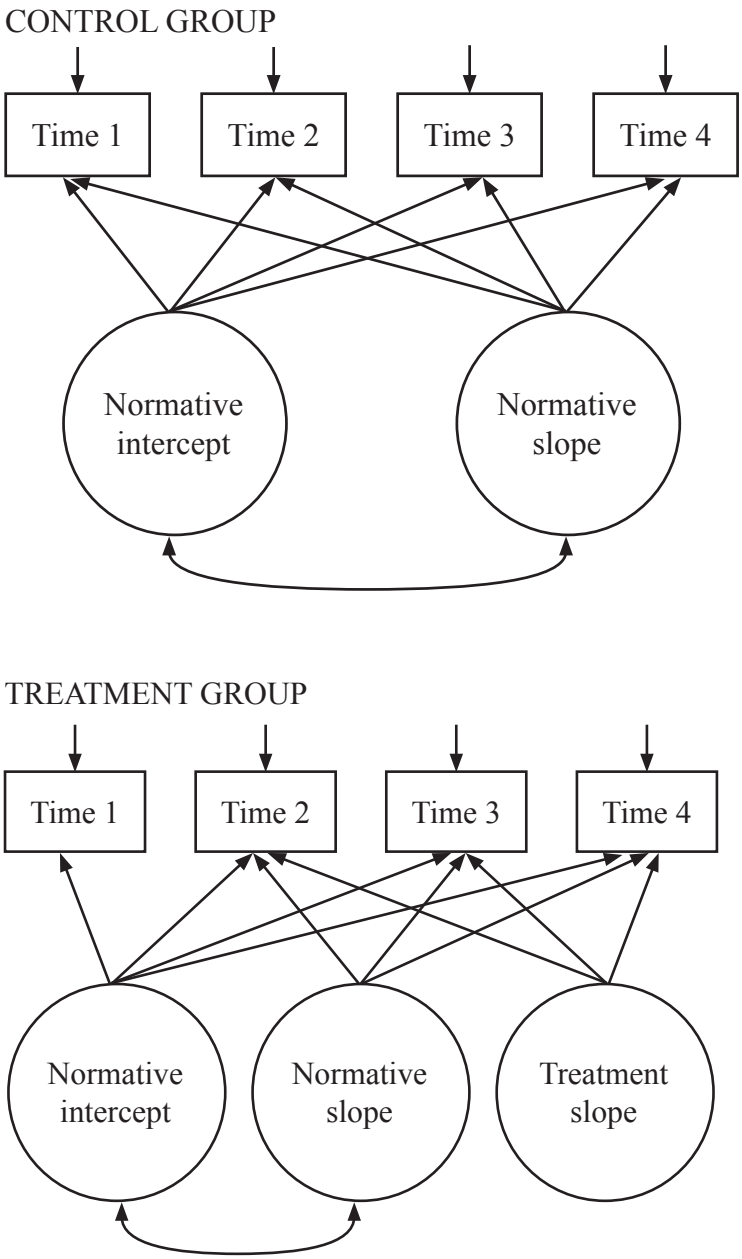

Figure 1. Path diagram of the latent curve model within the control group and within the treatment group.

Table 5

Means and Standard Deviations for Reading Outcomes

		Fall, 9th grade			Spring, 9th grade			Fall, 10th grade			Spring, 10th grade		
		n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Gates vocabulary SS	Comparison	177	80.07	8.25	160	79.89	9.40	158	81.34	8.17	140	83.20	7.91
	Treatment	170	80.65	8.43	158	79.89	10.16	148	81.66	8.25	124	84.60	8.46
Gates vocabulary ESS	Comparison	177	492.89	21.48	160	499.91	23.82	158	505.07	21.07	140	511.74	20.64
	Treatment	170	493.23	24.88	158	499.82	25.99	148	506.28	20.91	124	515.56	19.60
Gates comprehension SS	Comparison	178	79.02	10.00	160	83.01	10.43	152	82.92	7.51	140	86.66	8.94
	Treatment	169	81.22	9.85	158	82.42	10.25	140	83.53	8.86	124	87.54	8.63
Gates comprehension ESS	Comparison	178	491.43	25.83	160	506.04	26.49	158	512.84	20.04	140	524.87	21.83
	Treatment	169	496.24	26.89	158	503.47	28.62	148	513.84	22.17	124	527.58	18.99
TOSREC SS	Comparison	177	76.64	13.74	157	79.51	11.11	155	74.56	13.11	140	77.71	12.53
	Treatment	170	75.34	13.92	154	80.51	10.84	144	74.34	12.44	121	78.88	12.35
TOWRE-2 SW SS	Comparison	177	91.83	11.36	159	93.53	11.99	158	93.10	11.28	140	94.49	11.83
	Treatment	170	88.94	11.12	158	90.34	11.74	148	90.01	10.73	124	91.45	12.29
TOWRE-2 PD SS	Comparison	176	96.47	11.61	160	96.91	12.65	158	97.06	11.64	140	98.11	11.93
	Treatment	170	94.11	13.20	158	95.35	12.51	148	94.47	12.45	124	96.23	13.56
Vocabulary RS	Comparison	178	3.69	2.44	160	4.21	2.79	158	4.37	2.75	140	5.26	3.93
	Treatment	170	4.03	2.51	158	5.09	3.26	148	5.12	3.74	124	6.36	4.60

Note. Gates = Gates-MacGinitie Reading Tests; SS = Standard Score; ESS = Extended Scale Score TOSREC = Test of Silent Reading Efficiency and Comprehension; TOWRE-2 = Test of Word Reading Efficiency–Second Edition; SW = sight words; PD = phonemic decoding; RS = Raw Score.

Table 6

Parameter Estimates and Model Fit Indices From the Latent Growth Model

	TOSREC	Gates comprehension	Gates vocabulary	TOWRE-2 SW	TOWRE-2 PD	RIA vocabulary
Model parameters						
Means						
Intercept (control)	0	0	0	0	0	0
Intercept (treatment)	-1.57	4.96	0	-3.04***	-2.42	0
Normative slope	2.31**	32.96***	19.94***	2.8***	1.83	1.36***
Treatment slope	2.52*	-2.49	0.01	-0.22	0.69	1.00*
Effect size (Hedges' *g*)	0.18	-0.09	0.00	-0.02	0.06	0.41
Variances						
Intercept	85.05***	439.52***	295.95***	108.46***	116.04***	2.49***
Normative slope	5.3	249.12*	156.02	27.95*	0	3.98
Regression coefficient						
Slope on intercept	-0.15	0.05	-0.01	-0.02	0.155	0.83***
Regions of significance						
Pretest -2 *SD*	5.35***	-4.74	0.29	-0.13	-2.67	-1.85
Pretest -1.5 *SD*	4.64***	-4.18	0.17	-0.04	-1.83	-1.14
Pretest -1 *SD*	3.93***	-3.61	0.12	0.05	-0.99	-0.43
Pretest -.5 *SD*	3.23***	-3.05	0.07	0.14	-0.15	0.29
Pretest at the mean	2.52*	-2.49	0.01	-0.22	0.69	1.00*
Pretest +.5 *SD*	1.81	-1.93	0.10	-0.31	1.53	1.72***
Pretest +1 *SD*	1.10	-1.36	-0.10	-0.39	2.38	2.43***

	TOSREC	Gates comprehension	Gates vocabulary	TOWRE-2 SW	TOWRE-2 PD	RIA vocabulary
Pretest +1.5 *SD*	0.39	-0.80	-0.15	-0.48	3.22	3.15***
Pretest +2 *SD*	-0.31	-0.24	-0.21	-0.57	4.06	3.86***
Model fit						
$\chi 2$ (*df*)	4.98 (9)	19.40(8)	12.209 (8)	3.37 (8)	20.715(10)	10.93 (9)
CFI	1.00	0.93	0.98	1.00	0.98	0.99
TLI	1.00	0.90	0.98	1.00	0.98	0.99
RMSEA	.00	.09	.06	.00	.08	.04
(90% CI)	(.00-.05)	(.04-.14)	(.00-.11)	(.00-.04)	(.03-.13)	(.00 - .10)

Note. TOSREC = Test of Silent Reading Efficiency and Comprehension; Gates = Gates-MacGinitie Reading Tests; SW = sight words; TOWRE-2 = Test of Word Reading Efficiency–Second Edition; PD = phonemic decoding.

* $p < .05$. ** $p < .01$. *** $p < .001$.